Modality-Aware Discriminative Fusion Network for Integrated Analysis of Brain Imaging Genomics

Xiaoqi Sheng, Hongmin Cai, *Senior Member, IEEE*, Yongwei Nie, *Member, IEEE*, Shengfeng He, *Senior Member, IEEE*, Yiu-Ming Cheung, *Fellow, IEEE*, and Jiazhou Chen, *Member, IEEE*

***Abstract*— Mild cognitive impairment (MCI) represents an early stage of Alzheimer's disease (AD), characterized by subtle clinical symptoms that pose challenges for accurate diagnosis. The quest for the identification of MCI individuals has highlighted the importance of comprehending the underlying mechanisms of disease causation. Integrated analysis of brain imaging and genomics offers a promising avenue for predicting MCI risk before clinical symptom onset. However, most existing methods face challenges in: 1) mining the brain network-specific topological structure and addressing the single nucleotide polymorphisms (SNPs)-related noise contamination and 2) extracting the discriminative properties of brain imaging genomics, resulting in limited accuracy for MCI diagnosis. To this end, a modality-aware discriminative fusion network (MA-DFN) is proposed to integrate the complementary information from brain imaging genomics to diagnose MCI. Specifically, we first design two modality-specific feature extraction modules: the graph convolutional network with edge-augmented self-attention module (GCN-EASA) and the deep adversarial denoising autoencoder module (DAD-AE), to capture the topological structure of brain networks and the intrinsic distribution of SNPs. Subsequently, a discriminative-enhanced fusion network with correlation regularization module (DFN-CorrReg) is employed to enhance inter-modal consistency and between-class discrimination in brain imaging and genomics. Compared to other state-of-the-art approaches, MA-DFN not only exhibits superior performance in stratifying cognitive normal (CN) and MCI individuals but also identifies disease-related brain regions and risk SNPs locus, which hold potential as putative biomarkers for MCI diagnosis.**

***Index Terms*— Brain imaging genomics, early diagnosis, mild cognitive impairment (MCI), multimodal fusion.**

Manuscript received 16 January 2024; revised 15 May 2024; accepted 29 July 2024. Date of publication 23 August 2024; date of current version 5 May 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0112200; in part by the Key Research and Development Program of Guangzhou under Grant 2023B01J0002; in part by the National Natural Science Foundation of China under Grant 62325204, Grant U21A20520, Grant 62102153, Grant 62272326, and Grant 62172112; in part by the Science and Technology Project of Guangdong Province under Grant 2022A0505050014; in part by the Key-Area Research and Development Program of Guangzhou City under Grant 202206030009; in part by the Natural Science Foundation of Guangdong Province of China under Grant 2022A1515011162 and Grant 2023A1515012894; in part by Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2023B1515020097; in part by Singapore MOE Tier 1 Funds under Grant MSS23C002; in part by NRF Singapore under AI Singapore Program under Grant AISG3-GV-2023-011; and in part by the Alzheimer's Disease Neuroimaging Initiative (ADNI). *(Corresponding author: Jiazhou Chen.)*

Xiaoqi Sheng, Hongmin Cai, and Yongwei Nie are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China.

Shengfeng He is with the School of Computing and Information Systems, Singapore Management University, Singapore 188065.

Yiu-Ming Cheung is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, SAR, China.

Jiazhou Chen is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China, and also with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, Guangdong 523083, China (e-mail: csjzchen@scut.edu.cn).

Digital Object Identifier 10.1109/TNNLS.2024.3439530

NOMENCLATURE

Symbols	Explanation
$x, \boldsymbol{x}, \boldsymbol{X}$	Scalar, vector, and matrix.
n	Number of brain network nodes.
x_{pq}	(pq)th element of the matrix $\boldsymbol{X}$.
k	Number of modalities.
d	Dimension of data.
ℓ	Network layer.
d_{ℓ_k}	Dimension of the kth modality in the ℓth layer.
$\boldsymbol{I}$	Identity matrix.
N	Number of total samples.
c	Number of classes.
N_i^k	Number of samples for modality k in ith class.
$\boldsymbol{x}_j^k$	jth sample in the kth modality.
$\boldsymbol{x}_{ij}^k$	jth sample from the kth modality in ith class.
μ, σ	Mean, variance.
tr	Trace operation.
$\boldsymbol{W}, \boldsymbol{b}$	Weight matrices, bias vectors.
$\boldsymbol{w}^k$	Column of $\boldsymbol{W}^k$.
$\|\cdot\|_F$	Frobenius norm.

I. INTRODUCTION

MILD cognitive impairment (MCI) is the prodromal stage of Alzheimer's disease (AD) and other types of dementia, which is characterized by the presence of objectively measurable cognitive deficits without fulfilling criteria for a diagnosis of dementia [1]. The etiology of MCI is multifaceted and cannot be ascribed to a single factor. Instead, it is likely the result of a combination of progressive processes [2]. Although MCI does not necessarily progress to dementia, people with MCI are more likely to experience cognitive and behavioral deficits resulting from abnormal alterations in the brain's connectome [3]. Studies have shown that about 15% of people with MCI develop AD each year, while a significant number of these patients have a stable disease course [4]. As such,

offering timely and appropriate intervention is considered a crucial approach to slow down or prevent the progression of MCI to AD. Unfortunately, the subtle manifestation of symptoms during the MCI phase often leads patients to miss the optimum intervention period, consequently increasing the risk of progression to AD [5]. Thus, accurate diagnosis of MCI is crucial in mitigating the progression of brain disease and identifying potential curative factors.

Numerous neuroscience studies have revealed that MCI exhibits alterations in brain network connectivity preceding the manifestation of clinical signs and symptoms. This observation has paved the way for predicting the risk of MCI progression [6], [7]. Moreover, emerging evidence suggests that genetic variations are the genetic underpinnings of brain network dysfunction, offering a new window to study and understand the neuropathological mechanism underlying MCI. Consequently, the integration of brain networks and genetic factors provides comprehensive understandings of brain structures, functionalities, and genetic foundations, thereby improving the accuracy, reliability, and specificity of MCI diagnosis [8], [9]. The rapid development of modern neuroimaging techniques, such as structural magnetic resonance imaging (sMRI) and diffusion tensor imaging (DTI), has enabled the construction of human brain connectivity networks (i.e., brain networks) [10], [11], [12]. Concurrently, genotype data, particularly single nucleotide polymorphisms (SNPs), are extensively employed to explore genetic variations associated with neurodegenerative diseases [8]. These high-dimensional and complex brain imaging genomics data present both opportunities and challenges in extracting intrinsic features from different modalities and efficiently integrating valuable complementary information to ultimately achieve accurate diagnosis of MCI. However, existing research has primarily focused on data dimensionality reduction or simply merging multiple data as input features for model training, resulting in insufficient fusion of information and substandard performance [13].

The integration of brain imaging genomics data currently involves two main stages: feature extraction and feature fusion. In the feature extraction stage, previous studies have predominantly employed straightforward extensions of unimodal methods. These approaches merge multimodal data into elongated vectors and apply traditional feature selection techniques [14], [15], [16] to identify important features for disease diagnosis. However, these algorithms overlook the presence of high-order information between features in multimodal data, such as the topological structure of brain networks and genetic regulatory networks. To address this limitation, graph convolutional neural networks (GCNs) have emerged as a natural approach to leverage intricate brain topology and genetic higher-order features [17], [18], providing technical support for understanding the complex neuropathological mechanisms underlying MCI [19], [20]. For example, existing GCNs-based approaches utilize unimodal feature representations from gene and brain network data, as well as interindividual correlations, to build multimodal graphs as inputs for subsequent diagnostic modeling [21]. Although significant performance improvements have been achieved compared to traditional feature extraction methods, these GCNs approaches often fail to effectively exploit the intrinsic information within each modality. The main reason is that the transmission of information between brain regions combines strong local connections with efficient long-range connections [22], but relying solely on GCNs tends to disregard the long-range dependencies between brain regions [23]. To address the long-range dependency limitations of GCNs, some studies have also applied the Transformer structure to the graph domain [24], [25], [26]. Typically, these methods utilize a designed position embedding strategy to integrate nodal position information into embeddings, facilitating self-attention calculations for specific graphs. However, such approaches tend toward overfitting when applied to imaging genomics data characterized by limited datasets. Additionally, brain networks and SNPs exhibit information imbalance and inconsistent feature representations, posing challenges for a single feature extraction network to derive effective fused features. Consequently, recent studies have proposed customizing a subnetwork for each modality to extract modality-specific features, followed by the concatenation of these embeddings to obtain a fusion representation [27], [28], [29], [30]. In the feature fusion stage, existing methods face challenges in capturing complementary information between brain networks and genetic factors due to limited interactions between modalities. Therefore, researchers have developed correlation analyses and nonlinear fusion methods to explore the features interacting with brain networks and genes [31], [32]. Although these fusion methods have yielded satisfactory performance in brain disease diagnosis, they often omit valuable diagnostic class labels, limiting their ability to discover patterns of interest. Recent works have enhanced models by integrating diagnostic labeling components, leveraging discriminative information to guide the model toward more flexible and accurate diagnosis. To utilize available labels, studies have introduced discriminant regularization or category alignment to model features among different modalities into a discriminative common subspace [33]. However, the learning schemes of these methods struggle to capture subtle changes induced by brain lesions, as they primarily focus on capturing global geometric properties of the data while ignoring inter-sample similarity.

The aforementioned analyses dictate that successful brain imaging genomics fusion approaches should not only effectively capture modality-specific intrinsic information but also derive between-class discrimination information. However, current approaches still encounter the following limitations.

1) Numerous multimodal fusion techniques often assume data-type uniformity across modalities solely for simplicity, neglecting each modality's unique structural qualities.
2) Existing feature extraction techniques demonstrate sensitivity to high SNP noise contamination and disregard distant dynamic brain network node interactions. As a result, these methods lack the necessary capacity to capture individual modality features effectively.
3) Current feature fusion methods remain deficient in simultaneously deriving inter-modal consistency and between-class discriminability.

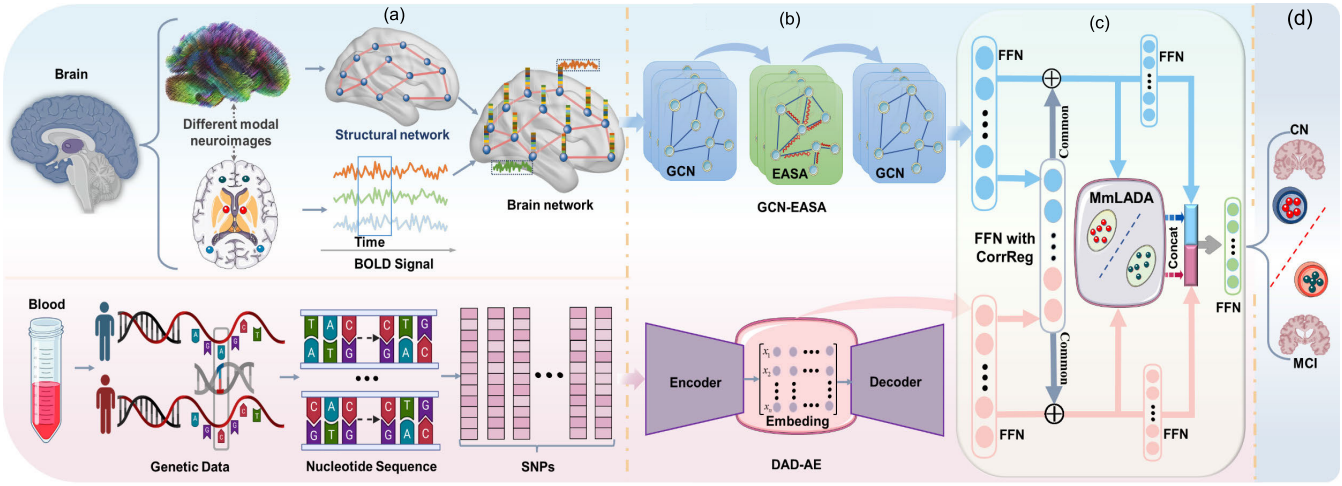


Fig. 1. Schematic of the proposed MA-DFN framework. (a) Brain network and genetic data undergo preprocessing before being utilized as the initial inputs for subsequent models. (b) Subsequently, two modality-specific feature extraction modules are proposed to learn low-dimensional representations of brain networks and genetic data, respectively. (c) Next, the learned low-dimensional representations are further extracted with discriminative and consistent features via the DFN-CorrReg module, where FFN is a feed-forward neural network. (d) Finally, the diagnosis of brain diseases relies on the extracted fusion features from different modalities.

To tackle these challenges, we propose a novel approach called modality-aware discriminative fusion network (MA-DFN). The MA-DFN incorporates two modality-specific feature extraction modules, which preserve each modality's structure and distribution. It also employs a robust fusion strategy to derive inter-modal consistency and between-class discriminability essential for MCI diagnosis. The main schematic of our proposed framework is illustrated in Fig. 1. In Fig. 1(a), the brain network is formed using the structural network inferred from sMRI and DTI data, complemented by node attributes extracted from the blood-oxygen-level-dependent (BOLD) signal acquired through functional magnetic resonance imaging (fMRI) scans. Moreover, the SNP data is derived from genome-wide genotype data. Subsequently, to extract local features and capture long-range dependency information from the structural-functional brain network, we design a graph convolutional network with an edge-augmented self-attention module (GCN-EASA), as shown in the first row of Fig. 1(b). Furthermore, we propose a deep adversarial denoising autoencoder module (DAD-AE) to capture the intrinsic distribution of SNPs and mitigate the impact of random noise on identifying important gene loci, as depicted in the second row of Fig. 1(b). Finally, we employ a discriminative-enhanced fusion network with a correlation regularization module (DFN-CorrReg) to eliminate inter-modal discrepancies using CorrReg. This module also learns modality-aware discriminative representations related to diagnostic labels, while preserving inter-modal similarity features within classes through multimodal local adaptive discriminant analysis (MmLADA), as depicted in Fig. 1(c). We evaluate the effectiveness of MA-DFN on the ADNI neuroimaging dataset. Extensive experimental results demonstrate that our proposed method achieves superior classification performance compared to state-of-the-art methods. Additionally, MA-DFN can identify regions of interest (ROIs) indicating brain lesions and risk gene loci, providing a novel perspective on understanding the pathogenesis of MCI. The contributions of our work are summarized as follows.

1) A GCN-EASA was designed to preserve the brain network topology and capture long-range dependency features between brain regions, facilitating deep interaction between structural and functional brain network features.
2) The proposed DAD-AE aims to effectively capture the intrinsic distribution of SNPs while minimizing the impact of random noise during the identification of crucial gene loci.
3) By tailoring the DFN-CorrReg module specifically for the fusion of brain network and SNPs, it effectively alleviates inconsistent feature interference in multimodal data while enhancing the clarity of decision boundaries between samples.
4) Integration of modality-specific and fusion modules into a unified model MA-DFN with collaborative optimization to mutually benefit. Extensive experimental results on real datasets have validated the effectiveness of MA-DFN in diagnosing brain diseases.

The remainder of this article is arranged as follows. Section II provides a summary of the theoretical background of related work that can support and offer insights into our rationale. Then, Section III introduces our proposed method accompanied by a theoretical analysis. Following that, Section IV presents the experimental results and analysis of the proposed method on the ADNI dataset. Finally, the further discussion and conclusion are given in Sections V and VI.

II. Preliminaries

This section primarily introduces essential and pertinent preliminary studies. To facilitate comprehension, Nomenclature summarizes the notation used in this paper.

A. Graph Convolutional Networks for Brain Networks

The self-organizing patterns of a brain network constructed by sMRI and DTI data can be depicted by a graph structure,

which can be represented visually as a weighted graph $G = (\mathcal{V}, \mathcal{E}, \boldsymbol{A})$. The graph depicts ROIs in the brain as nodes $\mathcal{V} = \{v_p \mid p \in 1, \ldots, n\}$, while edges $\mathcal{E} = \{a_{p,q} \mid (v_p, v_q) \in V \times V\}$ represent connections between nodes and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ denotes strength of connections between nodes, forming a structural scaffold for information exchange. Specifically, the attributes of each node are represented by the row vector $\boldsymbol{h}_p \in R^{1 \times d}$. The feature matrix $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n] \in R^{n \times d}$ of the graph is calculated by the Pearson correlation coefficient of the BOLD signal. The input graphs in the model consist of individual-level graphs, where each graph has the same number of nodes.

For a given brain network with adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and the feature matrix $\boldsymbol{H}$, a single GCN layer [20] can be expressed as a two-step process: feature transformation ($\boldsymbol{Z}' = \boldsymbol{H}^{(\ell-1)}\boldsymbol{W}$) and feature aggregation ($\boldsymbol{H}^{(\ell)} = \tilde{\boldsymbol{A}}\boldsymbol{Z}'$). Therein, $\boldsymbol{Z}'$ is the output of the feature transformation operation. $\tilde{\boldsymbol{A}}$ is a normalized matrix of the undirected graph G with self-connected $\boldsymbol{I}_n$, i.e., $\tilde{\boldsymbol{A}} = \hat{\boldsymbol{D}}^{-(1/2)}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-(1/2)}$ with $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_n$. Therein, $\hat{\boldsymbol{D}}$ is the degree matrix of the correlation with $\hat{\boldsymbol{A}}$. One common method in brain network analysis is to stack multiple layers of GCNs and incorporate nonlinear activation functions $f_\rho(\cdot)$ between successive layers ℓ to extract the network's nonlinear features, which follows the propagation rules:

$$\boldsymbol{H}^{(\ell)} = f_\rho\left(\tilde{\boldsymbol{A}}\boldsymbol{H}^{\ell-1}\mathbf{W}^{\ell-1}\right). \tag{1}$$

Based on the preceding theoretical analyses, GCNs can efficiently integrate both anatomical and functional neuroimaging data, thereby facilitating the exploration of intricate structural features within brain networks.

B. Multimodal Correlation Learning

The CCA-based subspace learning method serves as a prominent algorithm for analyzing multimodality imaging genomics data. Its primary goal is to identify the best linear transforms $\boldsymbol{W}^k, k = 1, 2$ for both imaging features $\boldsymbol{X}^1 = [\boldsymbol{x}_1^1, \ldots, \boldsymbol{x}_N^1]$ and genetic features $\boldsymbol{X}^2 = [\boldsymbol{x}_1^2, \ldots, \boldsymbol{x}_N^2]$ to maximize the total correlation of the low-dimensional embeddings $\boldsymbol{W}^{k^T}\boldsymbol{X}^k$ from two modalities. The energy function is given by

$$\begin{aligned} &\max_{\boldsymbol{W}^1, \boldsymbol{W}^2} \frac{1}{N}\operatorname{tr}\left(\boldsymbol{W}^{1^T}\boldsymbol{X}^1\boldsymbol{X}^{2^T}\boldsymbol{W}^2\right) \\ &\text{s.t.} \quad \frac{1}{N}\boldsymbol{W}^{1^T}\boldsymbol{X}^1\boldsymbol{X}^{1^T}\boldsymbol{W}^1 = \frac{1}{N}\boldsymbol{W}^{2^T}\boldsymbol{X}^2\boldsymbol{X}^{2^T}\boldsymbol{W}^2 = \boldsymbol{I}. \end{aligned} \tag{2}$$

The data matrices $\boldsymbol{X}^k$ are assumed to be centered for simplicity. However, directly applying (2) to deep networks may impede the alignment of modality-specific patterns in the common space and the exploitation of end-to-end advantages, potentially leading to subpar performance of network models when processing unseen data. An alternative solution is to embed the CorrReg [34] in the network

$$\mathcal{L}_{\text{CorrReg}} := \text{Corr} \approx \frac{\sum_{j=1}^{N}\left(\boldsymbol{y}_j^1 - \mu^1\right)\left(\boldsymbol{y}_j^2 - \mu^2\right)}{\sqrt{\sigma^1\sigma^2 + \epsilon}} \tag{3}$$

where the scalar ϵ is included to ensure the stability of the calculations. Here, $\boldsymbol{y}_j^k = \boldsymbol{W}^{k^T}\boldsymbol{x}_j^k$, $\mu^k = (1/N)\sum_{j=1}^{N}\boldsymbol{y}_j^k$, and $\sigma^k = \sum_{j=1}^{N}(\boldsymbol{y}_j^k - \mu^k)^2$. For output weight matrix $\boldsymbol{W}^k$, (3) focuses on enhancing the correlation of features by learning independently each output neuron $\boldsymbol{w}_l^k$, where $\boldsymbol{w}_l^k$ denotes the l column of $\boldsymbol{W}^k$. Importantly, (3) can be seamlessly integrated into a DNN to create a correlation-regularized fusion layer optimized through gradient descent (refer to [34] for additional information). This refinement diminishes modality-specific noise patterns while amplifying feature correlation across modalities. By leveraging the advantages of CorrReg, we integrated it into the fusion module of the proposed MA-DFN method, as illustrated in Fig. 2. This incorporation bolsters neuronal aggregation and exploits the end-to-end capabilities of deep models to enhance coherence modeling between brain networks and SNPs.

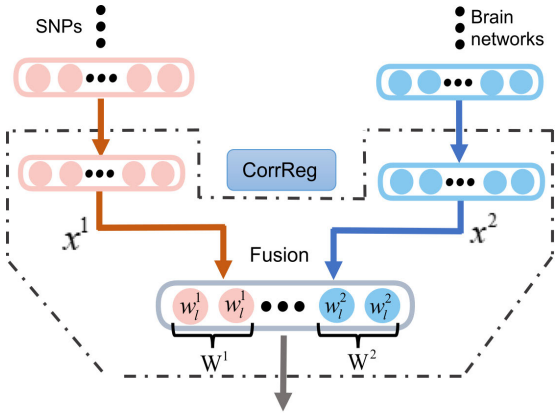


Fig. 2. Illustration of CorrReg fusion layer in extracting the common components of brain networks and SNPs.

III. Methodology

This section begins with a comprehensive overview of the modality-specific feature extraction modules, namely DAD-AE for SNPs and GCN-EASA for brain networks, followed by an exposition of the multimodal fusion module, DFN-CorrReg. Subsequently, the comprehensive framework and rationale behind the MA-DFN methodology are presented.

A. Modality-Specific Module

1) DAD-AE for SNPs: The presence of high-dimensionality and random noise in SNPs can lead to the occurrence of risk loci that are not associated with brain disease. Additionally, direct integration of genetic data with brain networks often results in an influx of noise rather than valuable features. Although current autoencoders can successfully achieve dimensionality reduction, they encounter difficulties in discerning irrelevant SNPs features for brain disease diagnosis. To tackle this challenge, a DAD-AE was developed with the goal of mitigating the adverse effects of SNPs-independent features on subsequent tasks and preserving the compressed data's distribution.

The DAD-AE comprises two key modules: the encoder and the decoder. The underlying concept behind the architecture involves a dimensionality reduction operation via an encoder to generate appropriate low-dimensional features of SNPs, followed by a dimensionality promotion utilizing a decoder to maintain the distribution of SNPs. During the training phase

of the DAD-AE, adversarial learning theory [35] is employed to reduce the sensitivity of the deep network to perturbations in SNP data. The DAD-AE takes a vector $\boldsymbol{X}$ as input and corrupts $\boldsymbol{X}$ to $\tilde{\boldsymbol{X}}$ with a worst case perturbation, represented as follows:

$$\tilde{\boldsymbol{X}} = \boldsymbol{X} + \alpha \cdot \text{sign}(\nabla_{\boldsymbol{X}} \mathcal{L}_C(f_\Theta(\boldsymbol{X}), \mathbf{y})) \tag{4}$$

where α is a learning rate value within the range of [0,1]. The sign represents the symbolic function. The f_Θ is a encoder network classifier. The $\mathbf{y}$ is the label of $\boldsymbol{X}$. The $\nabla_{\boldsymbol{X}} \mathcal{L}_C$ denotes the partial derivative of $\mathcal{L}_C$ with respect to $\boldsymbol{X}$, where $\mathcal{L}_C$ is cross-entropy loss (CEL) function. The encoder takes $\tilde{\boldsymbol{X}}$ as input and generates a low-dimensional representation $\boldsymbol{h} = s_e(\boldsymbol{b} + \boldsymbol{W}\tilde{\boldsymbol{X}})$, which is subsequently reconstructed by the decoder as $\boldsymbol{X}' = s_d(\boldsymbol{b}' + \boldsymbol{W}'\boldsymbol{h})$. Therein, s_e and s_d are the encoder and decoder activation functions. The DAD-AE parameters are optimized via mean square error (mse) back-propagation, minimizing the difference between the output $\boldsymbol{X}'$ and the input $\boldsymbol{X}$, as formulated in the following equation:

$$\mathcal{L}_{\text{mse}} = \frac{1}{N} \sum_{i=1}^{N} \left\| \boldsymbol{x}_i - \boldsymbol{x}'_i \right\|_2^2. \tag{5}$$

The low-dimensional representation $\boldsymbol{h}$ obtained by DAD-AE provides a clear and crucial feature representation of SNPs while preserving data distribution. Furthermore, this representation simplifies the fusion of features in subsequent stages.

2) GCN-EASA for Brain Networks: As stated in Section II-A, GCN provides a methodical approach to investigating differences in topological properties between brain networks of individuals diagnosed with MCI from elderly normal controls. This method has demonstrated its ability to capture local node features within brain networks, rather than long-range whole-brain-dependent features. Although GCNs can alleviate the problem of capturing long-range dependencies of brain networks by stacking deeper layers, this can lead to over-smoothing problems that render the model invalid. To tackle the above issues, a GCN-EASA module is proposed.

The GCN-EASA module retains GCN's ability to identify critical brain regions implicated in diseases by leveraging local neighborhood information, while extending the original transformer framework [36] via edge-augmented self-attention module (EASA). This augmentation allows identifying globally important brain regions within individual brain networks. The EASA contains node-embedding and edge-embedding channels, permitting edge information integration to explore long-range dependent network features. In the node-embedding channel, edge information is allowed to contribute to the node aggregation process [24]. This process can be represented as follows (as shown in Fig. 3):

$$\boldsymbol{h}_p^\ell = O_h^\ell \|_{\hbar=1}^{\mathbb{H}} \left(\sum_{q \in \mathcal{N}_i} \boldsymbol{w}_{pq}^{\hbar,\ell} \left(\boldsymbol{V}^{\hbar,\ell} \boldsymbol{h}_q^\ell \right) \right) + \boldsymbol{h}_p^\ell$$

$$\text{where, } \boldsymbol{w}^{\hbar,\ell} = \text{softmax}\left(\frac{\boldsymbol{Q}^{\hbar,\ell} \left(\boldsymbol{K}^{\hbar,\ell} \right)^T}{\sqrt{d_\ell}} + \boldsymbol{E}^{\hbar,\ell} \right) \tag{6}$$

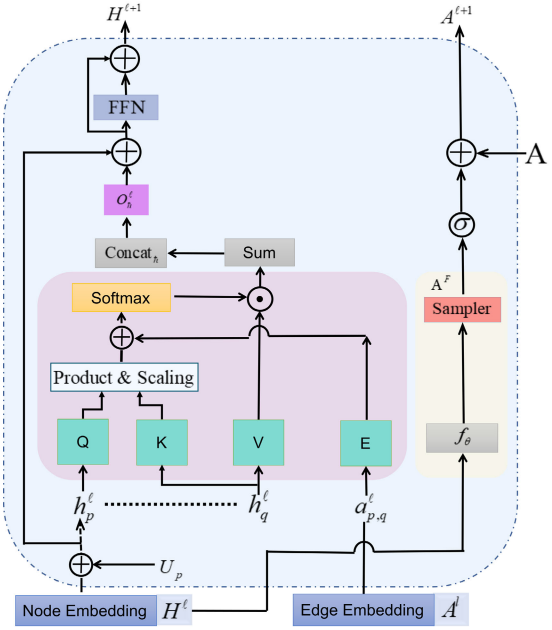


Fig. 3. Illustration of GCN-EASA in capturing information about long-range dependencies of brain networks and structure-function network interactions. The ⊙ denotes element-wise product.

where the ℓ is the number of layers. The $\hbar = 1$ to $\mathbb{H}$ denotes the number of attention heads, and $\|$ denotes concatenation. The $\boldsymbol{h}_p^\ell \in \mathbb{R}^{d_h}$ (for $1 \le p \le n$) is the input nodes embedding, where d_h is the node embeddings dimensionality. The $\hat{\boldsymbol{h}}_p^\ell$ is the attentional output. The $\boldsymbol{w}^{\hbar,\ell}$ is the attention matrix. The $O_h^\ell \in \mathbb{R}^{d_h \times \mathbb{H}n}$ is the learned output projection matrices. The $\boldsymbol{Q}^{\hbar,\ell}$, $\boldsymbol{K}^{\hbar,\ell}$, and $\boldsymbol{V}^{\hbar,\ell} \in \mathbb{R}^{n \times d_h}$ are key, query, and values generated through linear transformations of the embedding. The $\boldsymbol{E}^{\hbar,\ell} \in \mathbb{R}^{n \times n}$ is a concatenation of the learned linear transformed edge embeddings. In (6), the bias term $\boldsymbol{E}^{\hbar,\ell}$ is introduced to enhance the impact of edge embedding on attentional processes. This mechanism promotes a deep interaction between structural and functional brain networks. Furthermore, the multihead self-attention mechanism of GCN-EASA allows for unconstrained long-range dynamic interactions between brain regions, effectively compensating for the limitation of GCN's local convolutional aggregation. Subsequently, the attention output $\hat{\boldsymbol{h}}_p^\ell$ is fed into the feed-forward neural network (FFN) while being combined with the residual connection to yield the final output $\boldsymbol{H}^{\ell+1}$. The positional encoding of attention is represented by the d smallest non-trivial feature vectors obtained from the factorization of the graph Laplacian matrix, i.e., $\tilde{\boldsymbol{A}} = \boldsymbol{U}^T \boldsymbol{\Lambda} \boldsymbol{U}$, where $\boldsymbol{\Lambda}, \boldsymbol{U}$ are the eigenvalues and eigenvectors, respectively. Thus, the positional encoding of node p is defined as $\boldsymbol{U}_p[1:d]$ [24].

Moreover, the GCN-EASA module integrates an edge enhancement mechanism, enabling the learning of appropriate aggregation patterns during training rather than being confined to predetermined patterns. This mechanism indirectly improves the subsequent aggregation operation of the GCN and enhances the attentional mechanism's capability to capture crucial brain regions globally. Specifically, a set of auxiliary embeddings $\{\boldsymbol{h}_1'^\ell, \ldots, \boldsymbol{h}_n'^\ell\}$ is obtained using node embeddings after FFN. In this context, the pairwise functional profiles

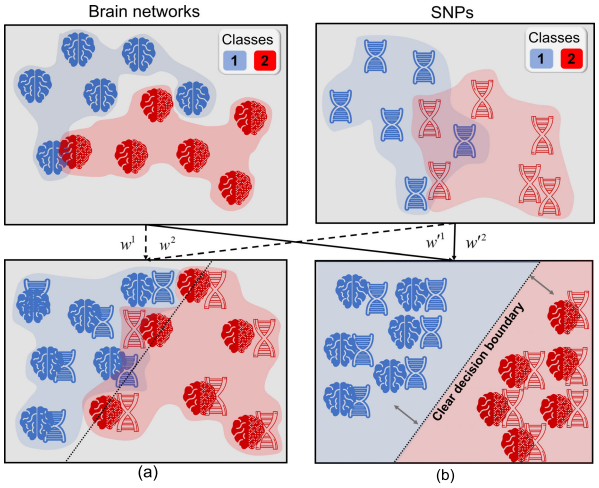


Fig. 4. Illustration of DFN-CorrReg in enhancing inter-modal consistency and between-class discrimination of brain networks and SNPs. (a) Inter-modal consistency. (b) Proposed DFN-CorrReg.

between brain regions p and q are defined as $\boldsymbol{A}^F := (a^{\ell}_{p,q}) = sigmoid(e^{-\|\boldsymbol{h}'^{\ell}_p - \boldsymbol{h}'^{\ell}_q\|_2^2})$. Despite potential imperfections in the initial connections of brain structure, they still cover a significant number of valuable actual topologies. Thus, the structural brain network and functional profiles are integrated to construct the edge of task-related brain regions, represented by the $\boldsymbol{A}^{\ell+1} = \tilde{\boldsymbol{A}} + \vartheta \boldsymbol{A}^F$. Therein, the ϑ is a non-negative equilibrium parameter that is iteratively updated during the model's learning phase.

B. DFN-CorrReg Module

Degenerative disorders are often manifest localized effects in specific regions of the brain, and these alterations are significantly associated with genetic variations [37]. One major drawback of existing algorithms is their indiscriminate integration of extracted brain network and genetic features, without distinguishing between specialized and general characteristics. This lack of differentiation interferes with the decision-making process of the classifier. To tackle this issue, we propose a DFN-CorrReg module specifically designed for integrating brain imaging genomics data. This module consists of two components, namely CorrReg and MmLADA. The purpose of CorrReg, as introduced in Section II-B, is to map brain networks and SNPs into a common space to extract their correlation information (inter-modal consistency), which is related to the occurrence and development of degenerative disorders, as shown in Fig. 4(a). Next, the MmLADA is proposed to address the under-utilization of labels and local information in CorrReg, thereby enhancing the effectiveness and discriminative capability in diagnosing brain diseases (between-class discrimination), as shown in Fig. 4(b). This section will focus on a detailed description of the proposed MmLADA.

The samples from the kth modality, denoted as $\mathcal{X}^k = \{\boldsymbol{x}^k_{ij} \in \mathbb{R}^{d_{\ell_k}} \mid i = 1, \ldots, c; j = 1, \ldots, N^k_i; k = 1, \ldots, v\}$. The proposed method, MmLADA, seeks to learn a set of linear transformations $\boldsymbol{W}' = [(\mathbf{w}'^1)^T, \ldots, (\mathbf{w}'^v)^T]^T$ that can project the feature representations obtained in the modality-specific stage onto a common discriminant space. Meanwhile, MmLADA automatically assigns different weights to each sample without additional parameters and eliminates inter-modal differences, enabling cross-modal feature interactions. Through this process, deep multimodal features projected into a common space are expected to exhibit compact within-class and inter-modality distributions, while displaying scattered between-class distributions. Compared to the method proposed in [38], MmLADA utilizes paired modal information to maximize between-class differences. Mathematically, the objective function is formulated as follows:

$$\min_{\mathbf{w}'^1, \ldots, \mathbf{w}'^v} \frac{\operatorname{tr}(\tilde{\boldsymbol{S}}_W) + \beta_1 \sum_{k=1}^{v} \|\mathbf{w}'^k\|_F^2}{\operatorname{tr}(\tilde{\boldsymbol{S}}_{\mathrm{T}}) + \beta_2 \operatorname{tr}(\tilde{\boldsymbol{S}}_M)}$$
$$\text{s.t.} \sum_{l=1}^{N^k_i} s_{ijl} = 1, s_{ijl} \geq 0; \quad j, l = 1, 2, \ldots, N^k_i \tag{7}$$

where

$$\tilde{\boldsymbol{S}}_W = \sum_{i=1}^{c} \sum_{k=1}^{v} N_i \sum_{j=1}^{N^k_i} \sum_{l=1}^{N^k_i} {s_{ijl}}^2 \left(\boldsymbol{x}^k_{ij} - \boldsymbol{x}^k_{il}\right)\left(\boldsymbol{x}^k_{ij} - \boldsymbol{x}^k_{il}\right)^T \tag{8}$$

$$\tilde{\boldsymbol{S}}_{\mathrm{T}} = \frac{1}{N} \sum_{k=1}^{v} \sum_{j=1}^{N^k} \sum_{l=1}^{N^k} \left(\boldsymbol{x}^k_j - \boldsymbol{x}^k_l\right)\left(\boldsymbol{x}^k_j - \boldsymbol{x}^k_l\right)^T \tag{9}$$

$$\tilde{\boldsymbol{S}}_M = \sum_{i}^{c} \sum_{k=m}^{v} \sum_{j=1}^{N^k_i} \sum_{l=1}^{N^k_i} \left(\boldsymbol{x}^k_{ij} \boldsymbol{x}^{mT}_{il}\right) \tag{10}$$

where $\tilde{\boldsymbol{S}}_{\mathrm{T}}$ and $\tilde{\boldsymbol{S}}_W$ denote the total- and within-class scatter matrices, and the matrix $\tilde{\boldsymbol{S}}_M$ represents the pairwise correlation in the same modality of the within-class data. The $N_i = \sum_{k=1}^{v} N^k_i$ represents the total number of samples belonging to the ith class across all modalities. In (7), the $\|\mathbf{w}'^k\|_F^2$ denotes Tikhonov regularization, which is employed to enhance the generalizability of the solutions. The s_{ijl} means the weight between the jth and lth samples in class i. The constraint on the s is imposed to prevent the occurrence of a trivial solution. The β_1 and β_2 are a non-negative parameter.

Formally, the $\tilde{\boldsymbol{S}}_W$ in the common space can be reformulated as follows [33]:

$$\tilde{S}_W = \left[\left(\mathbf{w}'^1\right)^{\top} \left(\mathbf{w}'^2\right)^{\top} \cdots \left(\mathbf{w}'^v\right)^{\top}\right] \begin{pmatrix} \boldsymbol{S}^{11}_{\boldsymbol{W}} & \cdots & \boldsymbol{S}^{1v}_{\boldsymbol{W}} \\ \vdots & \vdots & \vdots \\ \boldsymbol{S}^{v1}_{\boldsymbol{W}} & \cdots & \boldsymbol{S}^{vv}_{\boldsymbol{W}} \end{pmatrix} \begin{bmatrix} \mathbf{w}'^1 \\ \mathbf{w}'^2 \\ \vdots \\ \mathbf{w}'^v \end{bmatrix}$$
$$= \boldsymbol{W}'^T \boldsymbol{S}_W \boldsymbol{W}' \tag{11}$$

where the (k, m)th sub-matrix of $\boldsymbol{S}_W$ is defined as

$$(\boldsymbol{S}_W)^{km} = \begin{cases} 2 \sum_{i=1}^{c} N_i \sum_{j=1}^{N^k_i} \sum_{l=1}^{N^m_i} {s_{ijl}}^2 \left(\boldsymbol{x}^k_{ij} \left(\boldsymbol{x}^m_{ij}\right)^T - \boldsymbol{x}^k_{ij} \left(\boldsymbol{x}^m_{il}\right)^T\right) & k = m, \\ -2\boldsymbol{x}^k_{ij} \left(\boldsymbol{x}^m_{il}\right)^T & \text{otherwise.} \end{cases} \tag{12}$$

Likewise, $\tilde{\boldsymbol{S}}_{\mathrm{T}}$ and $\tilde{\boldsymbol{S}}_M$ can be, respectively, rewritten as $\tilde{\boldsymbol{S}}_{\mathrm{T}} = \boldsymbol{W}'^T \boldsymbol{S}_{\mathrm{T}} \boldsymbol{W}'$ and $\tilde{S}_M = \boldsymbol{W}'^T \boldsymbol{S}_M \boldsymbol{W}'$, where $\boldsymbol{S}_{\mathrm{T}}$ and $\boldsymbol{S}_M$ are a

partitioned matrix. The (k, m)th sub-matrix of $\boldsymbol{S}_{\mathrm{T}}$ and $\boldsymbol{S}_M$ is given by

$$(\boldsymbol{S}_{\mathrm{T}})^{km} = \begin{cases} 2\left(\sum_{j=1}^{N^k} \boldsymbol{x}_j^k \left(\boldsymbol{x}_j^m\right)^T - \frac{1}{N}\sum_{j=1}^{N^k} \boldsymbol{x}_j^k \sum_{l=1}^{N^m} \left(\boldsymbol{x}_l^m\right)^T\right) & k = m, \\ -\frac{2}{N}\sum_{j=1}^{N^k} \boldsymbol{x}_j^k \sum_{l=1}^{N^m} \left(\boldsymbol{x}_l^m\right)^T & \text{otherwise} \end{cases} \tag{13}$$

$$(\boldsymbol{S}_M)^{km} = \begin{cases} \sum_{i=1}^{c}\sum_{j=1}^{N^k}\sum_{l=1}^{N^m} \boldsymbol{x}_{ij}^k \left(\boldsymbol{x}_{il}^m\right)^T & k = m, \\ \boldsymbol{0} & \text{otherwise.} \end{cases} \tag{14}$$

Due to the typical nonconvex nature of the aforementioned optimization problem, a closed-form solution for the general trace ratio problem does not exist. Thus, an alternate optimization strategy is employed to minimize the objective function.

When s_{ijl} is fixed, the problem in (7) is converted into

$$\max_{\mathbf{w}'^1,\ldots,\mathbf{w}'^v} \frac{\mathrm{tr}\left(\boldsymbol{W}'^T(\boldsymbol{S}_{\mathrm{T}} + \beta_2 \boldsymbol{S}_M)\boldsymbol{W}'\right)}{\mathrm{tr}\left(\boldsymbol{W}'^T(\boldsymbol{S}_W + \beta_1 \boldsymbol{I})\boldsymbol{W}'\right)}. \tag{15}$$

The trace ratio problem in (15) can be reformulated as a generalized eigenvalue decomposition (GED) problem [39], where the objective is to determine the eigenvalues $\{\phi_i | i = 1, \ldots, c-1\}$ and their corresponding eigenvectors $\{\omega_i | i = 1, \ldots, c-1\}$ that satisfy $(\boldsymbol{S}_W + \beta_1 \boldsymbol{I})^{-1}\omega_i = \phi_i(\boldsymbol{S}_{\mathrm{T}} + \beta_2 \boldsymbol{S}_M)\omega_i$. Specifically, the ϕ_i denotes the ith largest eigenvalue of the GED, with its corresponding eigenvector ω_i, which is the ith column vector of the matrix $\boldsymbol{W}'$. Moreover, the sets $\{\phi_i\}$ and $\{\omega_i\}$ are the eigenvalues and eigenvectors of $(\boldsymbol{S}_W + \beta_1 \boldsymbol{I})^{-1}(\boldsymbol{S}_{\mathrm{T}} + \beta_2 \boldsymbol{S}_M)$, respectively.

The weight matrix $\boldsymbol{W}'$ derived from the GED method was not utilized in the optimization process of the backpropagation neural network. To address this limitation, we formulated a novel loss function that leverages eigenvalues to consolidate the discriminant directions in the feature space. Building on the theory presented in [39], (7) can be rearranged as $\mathrm{tr}(\boldsymbol{W}'^{-1}(\boldsymbol{S}_W + \beta_1 \boldsymbol{I})^{-1}(\boldsymbol{S}_{\mathrm{T}} + \beta_2 \boldsymbol{S}_M)\boldsymbol{W}') = \mathrm{tr}((\boldsymbol{S}_W + \beta_1 \boldsymbol{I})^{-1}(\boldsymbol{S}_{\mathrm{T}} + \beta_2 \boldsymbol{S}_M))$, which is equivalent to $\{\phi_i\}$. However, directly using ϕ_i to guide network parameter updates can result in trivial solutions. Thus, a threshold is introduced to filter out larger eigenvalues and maximize the number of directions for each eigenvector. The corresponding discriminative loss (DL) function is defined as follows:

$$\mathcal{L}_{\mathrm{DL}} = -\frac{1}{t}\sum_{i=1}^{t} \phi_i$$
$$\text{s.t. } \{\phi_1, \ldots, \phi_t\} = \{\phi_i \mid \phi_i < \min\{\phi_1, \ldots, \phi_{c-1}\} + \epsilon\}. \tag{16}$$

From the formulation, the proposed model can utilize all $c-1$ available feature dimensions as much as possible, rather than exclusively maximizing the largest feature values. By leveraging the eigenvalue-based objective function, the discriminative information from all projection directions is maximally preserved for each model in the latent common space. Meanwhile, the objective function enables end-to-end training of the network model with backpropagation. This allows for maximizing the expressive capacity of the neural network model to fully represent between-class discrepancy when learning common representations across models.

Fixing $\boldsymbol{W}'$, we simplify (7) to the following form:

$$\min_{s} \sum_{i=1}^{c}\sum_{k=1}^{v}\sum_{j=1}^{N_i^k}\sum_{l=1}^{N_i^k} {s_{ijl}}^2 \left\| \boldsymbol{W}'^T\left(\boldsymbol{x}_{ij}^k - \boldsymbol{x}_{il}^k\right)\right\|_2^2$$
$$\text{s.t. } \sum_{l=1}^{N_i^k} s_{ijl} = 1, s_{ijl} \geq 0. \tag{17}$$

According to the optimization strategy proposed by [38], the optimal value of s_{ijl} can be computed as follows:

$$s_{ijl} = \frac{1}{\sum_{k=1}^{v}\left\| \boldsymbol{W}'^T\left(\boldsymbol{x}_{ij}^k - \boldsymbol{x}_{il}^k\right)\right\|_2^2} \times \left(\sum_{p=1}^{N_i^k} \frac{1}{\sum_{k=1}^{v}\left\| \boldsymbol{W}'^T\left(\boldsymbol{x}_{ij}^k - \boldsymbol{x}_{ip}^k\right)\right\|_2^2}\right)^{-1}. \tag{18}$$

By iteratively updating $\boldsymbol{W}'$ and s, MmLADA achieves adaptive learning of the local discriminative relationships within the desired data subspace. Furthermore, the DFN-CorrReg module seamlessly integrates the local discriminative features extracted by MmLADA with the global consistency ensured by CorrReg. This synergy effectively captures brain genetic salience features pertinent to the task, while alleviating the challenges associated with capturing brain microtopology and exploring cross-modal feature fusion.

C. Architecture of the Proposed Network

Fig. 1 illustrates the comprehensive pipeline of the proposed method, comprising two essential components: modality-specific feature extraction modules and the DFN-CorrReg fusion module. The former is responsible for identifying effective features within each modality. Specifically, GCN-EASA effectively extracts microstructural information from localized regions and captures crucial brain regions in the global brain network. By preserving the underlying brain topology, this approach enables the exploration of the impact of altered complex brain interactions on subsequent diagnosis. Moreover, the GCN-EASA incorporates a DAD-AE, which enhances its ability to filter redundant features in SNP data while preserving the original distributional information via the mean-squared error loss $\mathcal{L}_{\mathrm{mse}}$. The latter component of our proposed method is the DFN-CorrReg fusion module, guided by $\mathcal{L}_{\mathrm{CorrReg}}$ and $\mathcal{L}_{\mathrm{DL}}$, which transform the extracted modality-specific features into a common latent space. This module effectively mitigates inconsistent features within and between modalities, while enhancing between-class discrimination. As a result, the final objective function of MA-DFN, with the introduction of the classification loss $\mathcal{L}_C$, is as follows:

$$\mathcal{L} = \min_{\Theta} \mathcal{L}_C + \lambda_1 \mathcal{L}_{\mathrm{CorrReg}} + \lambda_2 \mathcal{L}_{DL} + \lambda_3 \mathcal{L}_{\mathrm{mse}} \tag{19}$$

where the $\Theta = \{\mathbf{W}, \mathbf{b}\}$ is the model's parameter. Equation (19) involves non-negative parameters, namely λ_1, λ_2, and λ_3,

which play a crucial role in balancing the influence of each loss term.

IV. Experiments

A. Data Collection and Preparation

The subjects used in the experiments were obtained from the AD neuroimaging initiative (ADNI) database (www.loni.usc.edu/ADNI), which upheld rigorous technical standards to ensure data homogeneity and had licensed our use of these data. To guarantee the credibility of the results, the data were retained as much as possible and the number of subjects in different categories was kept balanced. Concretely, a total of 380 subjects were included in the study, comprising 235 cognitive normal (CN) individuals (94 males, 141 females; mean age 74.10 $\pm$ 7.58 years), and 145 individuals with MCI (84 males, 61 females; mean age 74.49 $\pm$ 7.42 years). These subjects were selected from the original ADNI database.

B. Quality Control

This study undertook a joint analysis of brain network and genomics data derived from the ADNI dataset, focusing on four specific modalities: T1-weighted sMRI, resting-state functional MRI (rs-fMRI), and DTI scans, as well as SNPs.

For T1-weighted MRI, the field of view (FOV) is 240 $\times$ 256$\times$208 mm, and the voxel size is isotropic at 1.0 mm, TR = 2.3 s. The diffusion MRI data was collected in 54 gradient directions with FOV 232 $\times$ 232 $\times$ 160 mm, voxel size 2.0 mm isotropic, TR = 7.2 s, and TE = 56 ms. The rs-fMRI data included 197 vol, with a 220 $\times$ 220 $\times$ 163 mm FOV and voxel size was 3.3 mm isotropic, TR = 3 s, TE = 30 ms, and flip angle = 90°. Genetic data was sequenced using high-throughput sequencing, with the ADNI dataset providing sequencing file formats including variant call format (VCF) and binary alignment map (BAM). In addition, appropriate preprocessing of the data was performed at the outset to enhance data quality.

For the brain network data, a standard preprocessing step was taken. Skull-stripping was performed on three modalities, and T1 and rs-fMRI were registered to DTI space using FLIRT in the FMRIB software library (FSL) [40]. The rs-fMRI images underwent spatial smoothing, slice-time correction, temporal prewhitening, global drift removal, and bandpass filtering (0.01–0.1 Hz) using the FEAT command in FSL. The DTI images were eddy-corrected using FSL and fiber-tracked by MedINRIA [41]. T1-weighted images were registered to DTI space using FSL FLIRT and segmented using the FreeSurfer package [42]. The Destrieux Atlas [43] was then used to partition the cerebral cortex into 148 regions, which served as the basis for constructing the structural connectivity matrix $\boldsymbol{A}$. Weights were assigned based on the number of fibers connecting each pair of regions. Functional connectivity was estimated using the Pearson correlation coefficient between the BOLD time series obtained from two ROIs identified in rs-fMRI data.

In SNP data processing steps, we followed the canonical PLINK (version 1.90) [44] whole genome association analysis pipeline. Specifically, the raw whole-genome genotype data from the ADNI dataset was consolidated into a single dataset comprising 1 540 335 SNPs from 1877 subjects. During the subsequent quality control, any SNPs with a missing genotype rate exceeding 0.001 were removed, as were any SNPs with a minor allele frequency of less than 0.1. Additionally, SNPs that did not match the brain network data were also excluded. The Hardy–Weinberg equilibrium (HWE) checking steps filtered out any SNPs with significant deviations at a P-value of 0.001. Finally, across 380 subjects, the remaining 26 885 SNPs were used to compute a simple additive effect for each SNP locus based on its variant type, without any phenotype information involved. Each SNP value is coded as 0, 1, or 2 in an additive manner for the model inputs, indicating the number of minor alleles. Such data could be directly utilized to train deep learning models.

C. Comparative Methods and Settings

To evaluate the effectiveness of the MA-DFN model in classifying CN and MCI, we conducted a comparative experiment using the ADNI dataset. The comparison methods were categorized into four groups as follows.

1) *Baselines:* Linear discriminant analysis (LDA), Adaboost, and Lasso [14].
2) *Multimodal Correlation Analysis:* Canonical correlation analysis (CCA) [45], FDR-corrected sparse canonical correlation analysis (FDR-CSCCA) [16], sparse multimodal multitask learning (SMML) [46], improved multitask sparse CCA (IMTSCCA) [47].
3) *Non-geometric deep learning:* Stage-wise deep neural network (SWDNN) [48], genetic and multimodal imaging data using neural-network designs framework (G-MIND) [27], Grad-CAM-guided convolutional collaborative learning (GCAM-CCL) [15], deep canonically correlated sparse autoencoder (DCCS-AE) [49], deep collaborative learning (DCL) [50], deep generative-discriminative learning (DGDL) [29], supervised deep generalized CCA (SDGCCA) [51], imaging genetic deep neural network system (IGNET) [28], deep self-reconstruction sparse CCA (DS-SCCA) [52], and multiview LDA network (MLDA) [33].
4) *Geometric deep learning:* Brain imaging genetics by graph neural network (BIG-GRAPH) [17], edge-variational graph convolutional network (EV-GCN) [20], multimodality multiview graph representations and knowledge embedding (MMGK) [21], graph transformer (GTF) [24], and brain network transformer (BRAINNETTF) [25].

1) Data Settings and Metrics: In this experiment, the original dataset was divided into a training set comprising 80% of the data and a test set consisting of the remaining 20%. Subsequently, a fivefold cross-validation was conducted on the training set to evaluate both our method and the comparative methods. The trained model was then applied to the test dataset. The experiment was repeated independently ten times to ensure robustness. The experimental results were recorded to assess the statistical power of our method compared to other

methods. Five evaluation metrics were employed to measure the algorithm's performance, including accuracy (ACC), recall, precision, area under the ROC curve (AUC), and $F1$-score.

2) Model Settings: The proposed MA-DFN model in this study leverages data from two distinct modalities: brain networks and SNPs. To accommodate each modality, specialized designs are incorporated into the networks during their respective modules. The feature extraction module for the brain networks comprises two alternating layers of GCN and EASA, respectively. Similarly, the DAD-AE module is employed to extract low-dimensional features from the SNP data. This module consists of four FFN layers, with two dedicated to encoding and the remaining two focused on decoding. Additionally, the DFN-CorrReg module incorporates CorrReg and MmLADA, including five FFN layers. Ultimately, classification is executed using an FFN layer and the softmax function. The loss function optimizer is Adam. The model undergoes training for 100 epochs with an initial learning rate of $1e^{-3}$. Regularization is implemented by incorporating an initial weight decay of $1e^{-4}$. Hyperparameter selection is conducted automatically using the open-source toolkit Optuna. To ensure the fairness of the experiment, empirical adjustments were made to the parameters of both MA-DFN and the other methods used for comparison. Furthermore, a paired-sample T-test was conducted to compare our proposed method with the suboptimal comparison methods. The symbol $\star$ indicates that the proposed method's results have significant differences (p-value < 0.05) compared to the suboptimal algorithm.

D. Experimental Results in the ADNI Dataset

This section aims to evaluate the diagnostic performance of the proposed method in brain imaging genomics data. The experimental results for each method are summarized in Table I. From Table I, it can be observed that the proposed algorithm has superior performance in terms of accuracy, AUC, and $F1$-score compared to other methods. The striking results well demonstrate the usefulness of the proposed method, highlighting its potential in processing multimodal data involving complex brain networks and high-dimensional SNPs for diagnosing brain diseases. Additionally, the following conclusions can be drawn from Table I.

First, MA-DFN outperforms non-geometric deep learning methods in almost all classification metrics. Specifically, DS-SCCA is considered as the optimal geometric deep learning method in our experiments. However, our proposed MA-DFN significantly superior to DS-SCCA, achieving improvements of 5.8%, 8.2%, 3.8%, 5.7% and 6.5% in terms of accuracy, recall, precision, AUC, and $F1$-score, respectively. The consistency improvement arises from the separate consideration of brain networks and SNP data structure differences in the MA-DFN method, instead of treating all modalities as a uniform vector structure like non-geometric deep learning methods. Thus, non-geometric deep learning methods present challenges in their application to data with geometric characteristics, especially brain network data. These challenges can lead to the extraction of inaccurate features by the models, ultimately impacting the diagnosis of brain diseases. Contrastingly, MA-DFN designed two modules, namely GCN-EASA and DAD-AE, which exhibit excellent adaptability to brain networks and SNP data existing topological differences. This means that the feature representations learned by MA-DFN are not only more accurate but also possess geometrically meaningful properties.

Second, MA-DFN exceeds both EV-GCN and BIG-GRAPH by a large margin in all metrics except recall. For instance, in the CN versus MCI task, there were observed gaps of 3.8% and 2.9% in diagnostic accuracy. The primary reason is that the EV-GCN and BIG-GRAPH overlook the importance of long-range dependent features between brain regions and the interference of noise from SNPs. The EV-GCN and BIG-GRAPH assume that all SNP loci are beneficial for brain disease diagnosis and construct them into graph data, which is then combined with the brain network and fed into the GCNs model. However, the above algorithms only emphasize the ability of GCNs to preserve the topological characteristics of the data, ignoring the limitations of information transmission between nodes. Despite the ability of GTF and BRAINNETF to reduce the limitation of information transmission between long-range nodes of GCNs, all of them suffer from performance degradation due to the difficulty of effectively mining local and discriminative features of brain networks. Additionally, the direct integration of brain networks with SNP data may result in the issue of excessive noise mixing. As a result, the aggregated information from the nodes may not always be useful, impeding the performance of the GCNs model. In contrast, the MA-DFN model incorporating DAD-AE and GCN-EASA achieves the desired classification performance by effectively suppressing noise and fully exploiting both local and global features.

Third, the MA-DFN with MLDA method underperforms the MA-DFN with MmLADA method across all four metrics: accuracy, recall, AUC, and $F1$-scores. This discrepancy stems from MmLADA's ability to adaptively learn the local data structure in the subspace and improve the inter-modal similarity of within-classes. Additionally, the MA-DFN model with MmLADA outperforms the CCA-only method by a significant margin. The unsatisfactory performance of CCA-based method can be attributed to the challenge posed by the intra-modal heterogeneity and lack of between-class discriminability, as demonstrated in Fig. 4. Simply maximizing the inter-modal consistency feature is insufficient to effectively address this issue. In contrast, the MA-DFN employs the DFN-CorrReg module that incorporates CorrReg and MmLADA. This module not only maximizes the consistency inter-modal but also effectively resolves the issue of ambiguous decision boundaries. Furthermore, experimental results also demonstrate the strong performance of MA-DFN, highlighting its practical value.

E. Parameter Sensitivity

As stated in Section III-C, building a multimodal diagnostic model for brain diseases involves optimizing three hyperparameters, namely λ_1, λ_2, and λ_3. Given that λ_1 and λ_2 are involved in the modality fusion subproblem and λ_3 corresponds to the subproblem of reconstruction, these parameters can be analyzed in two groups. Specifically, we analyzed λ_1

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT CLASSIFICATION METHODS FOR CN VERSUS MCI

Model	Method	Data	Accuracy	Recall	Precision	AUC	F1-score
Shallow	LDA	FBN + SBN + SNPs	0.592±0.021	0.870±0.041	0.627±0.024	0.498±0.012	0.728±0.017
	Lasso		0.579±0.007	0.611±0.046	0.698±0.046	0.569±0.049	0.641±0.031
	Adaboost		0.605±0.002	0.928±0.023	0.634±0.046	0.506±0.036	0.747±0.003
	SMML		0.585±0.031	0.697±0.064	0.674±0.075	0.566±0.040	0.667±0.052
	CCA	FBN + SNPs	0.622±0.053	**0.965±0.071**	0.681±0.084	0.581±0.067	0.757±0.050
	IMTSCCA		0.625±0.053	0.963±0.022	0.657±0.060	0.506±0.036	0.747±0.041
	FDR-CSCCA		0.641±0.051	0.768±0.039	**0.974±0.015**	0.662±0.049	0.561±0.066
Deep	G-MIND	FBN + SBN + SNPs	0.736±0.026	0.623±0.025	0.827±0.098	0.735±0.026	0.699±0.025
	GCAM-CCL		0.741±0.030	0.639±0.142	0.804±0.086	0.729±0.018	0.698±0.046
	SWDNN		0.781±0.012	0.596±0.112	0.952±0.061	0.776±0.024	0.726±0.061
	DS-SCCA		0.784±0.042	0.660±0.062	0.879±0.010	0.784±0.042	0.753±0.034
	DGCCA		0.748±0.067	0.631±0.082	0.822±0.098	0.749±0.071	0.709±0.091
	EV-GCN		0.804±0.052	0.750±0.088	0.824±0.032	0.804±0.049	0.791±0.056
	MMGK		0.822±0.035	0.781±0.099	0.846±0.065	0.819±0.040	0.807±0.055
	GTF	FBN + SBN	0.772±0.027	0.699±0.092	0.827±0.081	0.771±0.029	0.751±0.051
	BRAINNETTF		0.809±0.044	0.839±0.069	0.837±0.036	0.805±0.046	0.801±0.077
	DCCS-AE	FBN + SNPs	0.711±0.042	0.571±0.058	0.692±0.073	0.712±0.029	0.664±0.033
	DCL		0.709±0.045	0.526±0.084	0.700±0.084	0.674±0.014	0.592±0.025
	DGDL		0.728±0.030	0.556±0.138	0.798±0.119	0.704±0.024	0.643±0.067
	IGNET		0.752±0.055	0.665±0.106	0.818±0.130	0.745±0.053	0.721±0.059
	BIG-GRAPH		0.813±0.040	0.797±0.083	0.835±0.026	0.813±0.038	0.797±0.045
Proposed method	MA-DFN (with MLDA)	FBN + SBN + SNPs	0.831±0.037	0.663±0.119	0.958±0.049	0.817±0.050	0.776±0.080
	MA-DFN (with MmLADA)		**0.842±0.023**	0.742±0.036	0.917±0.057	**0.841±0.017***	**0.818±0.020***

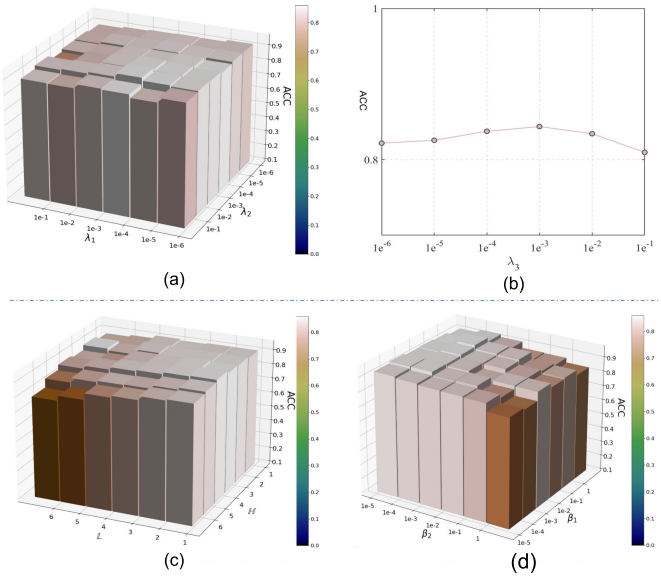


Fig. 5. Sensitivity analysis of the MA-DFN regarding (a) and (b) hyperparameters λ_1, λ_2, and λ_3, (c) number of attention heads $\mathbb{H}$ and layers $\mathbb{L}$, and (d) MmLADA parameters β_1 and β_2.

and λ_2 while keeping λ_3 fixed and analyzed λ_3 while keeping λ_1 and λ_2 fixed. This approach allowed us to observe the impact of these factors on the model's accuracy. Additionally, the diagnostic performance of the model is influenced by the number of attention heads ($\mathbb{H}$) and layers ($\mathbb{L}$) in GCN-EASA, as well as the parameters β_1 and β_2 of MmLADA. Therefore, this experiment focuses on examining the effects of these parameters on the final classification results. Specifically, two sets of CN versus MCI classification experiments were conducted using the ADNI dataset. These experiments utilized a grid search method to determine the optimal range of parameters within a predetermined range, which ultimately led to achieving the highest classification accuracy for the model. These results are reported in Fig. 5.

According to the experimental results presented in Fig. 5(a), MA-DFN exhibits a relatively stable accuracy score within a specific range. Additionally, Fig. 5(a) demonstrates that the model achieves higher classification accuracy with λ_1 ranging from $1e^{-4}$ to $1e^{-5}$ and λ_2 ranging from $1e^{-2}$ to $1e^{-3}$. Notably, the discriminative hyperparameter values λ_2 are larger than the consistent hyperparameter values λ_1. This suggests that the proposed MmLADA facilitates the model to better differentiate brain disease characteristics, while CorrReg ensures that the model captures commonalities. Based on the results presented in Fig. 5(b), it can be concluded that the model has good accuracy when the value of λ_3 is within the range of $[0.0001, 0.01]$. Fig. 5(c) illustrates the classification accuracy of MA-DFN on the CN versus MCI task while varying the values of $\mathbb{H}$ and $\mathbb{L}$. It is observed that using larger values for $\mathbb{L}$ introduces additional noise, leading to a decrease in the performance of the model. Setting a smaller value for $\mathbb{H}$ limits the model's expressiveness. We recommend setting $\mathbb{L}$ within the range of $[1, 2]$ and $\mathbb{H}$ within the range of $[2, 4]$. Fig. 5(d) reveals that the model performs relatively better when β_1 is in the range of $[0.001, 0.1]$ and β_2 is in the range of $[0.0001, 0.00001]$. In summary, the MA-DFN model has relatively stable ACC scores within a specified range.

F. Model Overfitting Strategies

One consensus in the deep learning community is that obtaining a good representation of the data for classification tasks via more nonlinear hidden layers is crucial. However,

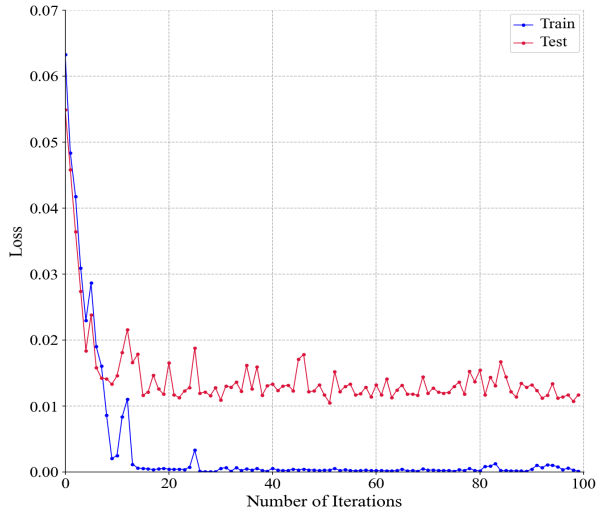


Fig. 6. Convergence curve of loss.

in the case of brain imaging genomics, the availability of a limited training sample size can lead to model being susceptible to overfitting. Although data augmentation has the potential to improve the classification accuracy of the model, it can also introduce unpredictable effects on brain networks and SNPs. Thus, this strategy was not included in the experiment. The primary focus of our efforts to prevent overfitting lies in the model's architecture and the training process. On the one hand, the model incorporates batch normalization, dropout, and early stopping strategies to mitigate the issue of overfitting. On the other hand, the model enhances robustness through the design of the DFN-CorrReg modules, enabling it to uncover essential features related to the classification task within the suitable hidden layers. Fig. 6 shows the convergence curves of both the training and test datasets. As depicted in Fig. 6, the loss of the test dataset exhibits no significant increase and maintains a steady-state gap with the training set. This observation indicates that despite limited training samples, the proposed method still exhibits strong generalization abilities.

G. Ablation Studies for MA-DFN

The proposed MA-DFN encompasses four innovative modules, namely GCN-EASA, DAD-AE, MmLADA, and CorrReg, with the aim of enhancing the performance of existing brain imaging genomics classifications. To evaluate the efficacy of each module in diagnosing MCI, we conducted an ablation experiment. The results of the experiment are displayed in Table II. The table uses ✓ to indicate that the module is utilized, while × denotes that it is not implemented.

Table II reveals that the proposed MA-DFN, which comprises four module combinations, significantly outperforms all other combinations across all metrics with respect to MCI classification. Furthermore, each module demonstrates enhanced performance in detecting MCI, which serves as evidence for the efficacy of our design. During the modality-specific module, the GCN-EASA and DAD-AE modules extract important features of brain networks and SNPs for classification tasks, resulting in a minimum of 1% improvement in accuracy when compared to models without these modules. Moreover,

TABLE II
CLASSIFICATION PERFORMANCE OF MA-DFN FOR ABLATION EXPERIMENT OF CN VERSUS MCI

Modules				Metrics		
EASA	DAD-AE	CorrReg	MmLADA	Accuracy	AUC	F1-score
×	×	×	×	0.804	0.803	0.802
×	✓	×	×	0.810	0.806	0.799
✓	✓	×	×	0.816	0.808	0.803
✓	✓	✓	×	0.827	0.834	0.813
✓	✓	×	✓	0.837	0.838	0.815
✓	✓	✓	✓	**0.842**	**0.841**	**0.818**

the DFN-CorrReg module demonstrates superior classification performance compared to a simple concatenation approach. This is primarily because the DFN-CorrReg module effectively improves the model's ability to capture discriminative features relevant to the diagnostic task by suppressing discrepant features between brain networks and genetic data. Furthermore, the seamless integration of the MmLADA and CorrReg modules of DFN-CorrReg into the model effectively addresses the challenges related to ambiguity and weak generalization in brain imaging genomics classification tasks.

H. Effectiveness of Multimodal Fusion

Integrating multimodal brain networks and genetic data has become an increasingly popular research field. By linking these modalities together, researchers can facilitate the discovery and accurate prediction of biomarkers associated with MCI. To verify the efficacy of the proposed method for fusing multimodal data, the ablation experiments were conducted on the ADNI dataset to analyze the impact of unimodal features and their combined fusion. In unimodal experiments, the MA-DFN method retained the modality-specific module relevant to the input modality and performed the disease diagnosis using the FFN in conjunction with the necessary loss function. Similarly, the SWDNN and G-MIND methods utilized similar settings to the MA-DFN algorithm in the unimodal task, whereas LDA and Adaboost did not require additional settings. The experimental outcomes are presented in Fig. 7.

Based on the classification results shown in Fig. 7, the proposed MA-DFN method surpasses unimodal data in all classification metrics when applied to multimodal data. This result confirms the essential role of multimodal data in the diagnosis of brain diseases and demonstrates the algorithm's effectiveness in dealing with heterogeneous multimodal data. Additionally, Fig. 7(a) and (b) illustrates that MA-DFN outperforms the comparison method in analyzing unimodal data. This finding implies that the modality-specific module of MA-DFN effectively learns alterations in brain network topology and diagnostic-related gene loci. Furthermore, the results in Fig. 7 demonstrate that shallow machine learning algorithms, such as LDA and Adaboost, struggle to mine meaningful features across modalities on multimodal data, highlighting the limitations of shallow machine learning algorithms. In contrast, deep learning approaches, like SWDNN and MA-DFN, outperform unimodal classification and achieve superior fusion results for multimodal data. Moreover, MA-DFN yields an

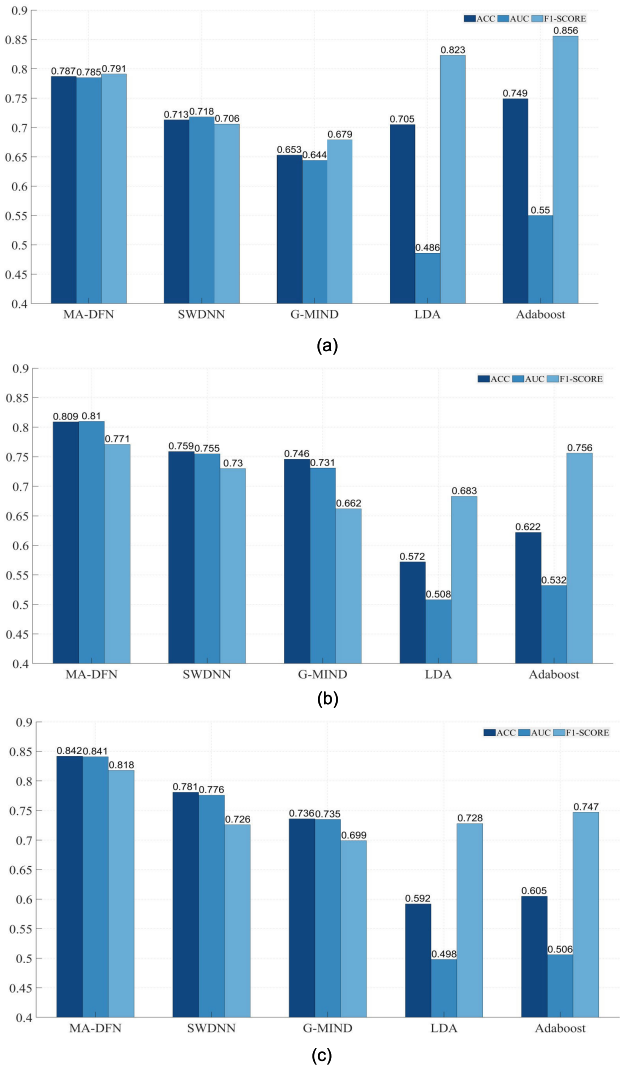
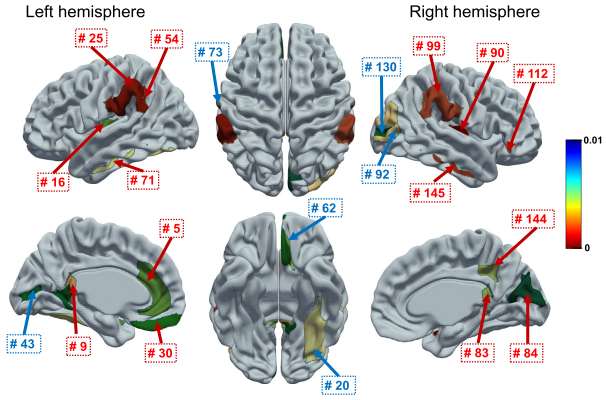


Fig. 7. Investigating the impact of multimodal dataset in brain imaging genomics on model performance. (a) Brain network dataset. (b) SNP dataset. (c) Brain network+SNP dataset.

accuracy score improvement of approximately 2.4% compared to shallow machine learning methods. Moreover, it was observed that SWDNN produced inferior classification results compared to our proposed MA-DFN. This phenomenon could potentially be attributed to the difficulty of SWDNN in eliminating modality-specific redundant features, which can lead to ineffective integration of multimodal features. Additionally, there may also be a lack of a suitable fusion mechanism to effectively integrate multimodal features, ultimately resulting in a decline in overall performance.

Upon reviewing Fig. 7(a)–(b) and (c) in conjunction, an intriguing observation was made. The MA-DFN algorithm [see Fig. 7(a) and (b)] that utilizes only unimodal data outperforms most of the multimodal algorithms [see Fig. 7(c)] in terms of ACC, AUC, and $F1$-score metrics. Two potential explanations emerge as prominent factors. A plausible explanation is that the MA-DFN is able to more efficiently capture modality-specific crucial information compared to other algorithms during the feature extraction phase. Another draws on the feature fusion viewpoint. A straightforward fusion of brain



Index	ROI Label	Index	ROI Label
1	# 90 (G_front_sup)	11	# 144 (S_suborbital)
2	# 145 (S_subparietal)	12	# 5 (G_and_S_transv_frontopol)
3	# 54 (S_front_sup)	13	# 16 (G_front_sup)
4	# 9 (G_cingul-Post-dorsal)	14	# 73 (S_temporal_sup)
5	# 99 (G_pariet_inf-Angular)	15	# 25 (G_pariet_inf-Angular)
6	# 92 (G_insular_short)	16	# 30 (G_precuneus)
7	# 20 (G_occipital_sup)	17	# 109 (G_temp_sup-Plan_polar)
8	# 71 (S_subparietal)	18	# 43 (Pole_temporal)
9	# 62 (S_orbital_lateral)	19	# 130 (S_intrapariet_and_P_trans)
10	# 83 (G_cingul-Post-dorsal)	20	# 84 (S_intrapariet_and_P_trans)

(b)

Fig. 8. Spatial locations of the top-20 significant brain regions. (a) CN-MCI significant regions. (b) List of top-20 significant brain regions from (a).

networks and SNPs is susceptible to noise and data heterogeneity. To address this issue, the introduction of MmLADA and CorrReg in the fusion stage significantly improved MCI classification outcomes. This outcome emphasizes the criticality of MA-DFN in effectively extracting features from each modality and aggregating them into a common latent subspace with both discriminative and consistent features.

V. Discussion

A. Discriminative Brain Regions

This section aims to assess the effectiveness of GCN-EASA in pinpointing discriminative brain regions. To achieve this goal, a systematic analysis of the brain information revealed by attention was performed at the group level. All the results presented in this section are derived from the testing dataset.

To examine the robust association between the critical brain regions identified by GCN-EASA and the progression of neurodegenerative diseases, a statistical multivariate t-test was employed to discern significant brain regions (p-value < 0.01) distinguishing CN from MCI. The top-20 significant brain regions identified by GCN-EASA were projected onto the cortical surface, as illustrated in Fig. 8(a). Fig. 8(b) displays the information regarding the significant brain regions (nodes), with the table being organized based on descending order of brain region significance. Therein, brain regions linked to default mode networks (DMNs) are highlighted in red, while those not associated are marked in blue.

Fig. 8 displays multiple regions within the frontal gyrus, specifically superior frontal sulcus [53]. The findings from these visualizations and other identified regions have been extensively documented in numerous ADNI reports. One notable observation is that the high significance observed in the

TABLE III
LIST OF TOP-15 RISK SNPS LOCUS

SNPs IDs	Allele	Gene	CHR
rs3851179	A	BINI	11
rs3764650	G	ABCA7	19
rs10498633	T	SLC24A4	14
rs574855	G	ANKRD60	20
rs6557170	A	ESR1	6
rs439401	T	APOE	19
rs405509	C	APOE	19
rs2491133	T	SDC3	1
rs749008	C	BINI	2
rs727153	G	LOC124900169	4
rs2829973	G	APP	21
rs3752193	T	OR7A10	19
rs1801133	T	MTHFR	1
rs6685309	C	FGGY	1
rs7081208	A	FRMD4A	10

postdorsal, angular [54], and subparietal [55] regions, aligning with current findings in AD diagnosis. Specifically, subparietal regions are known to have a critical role in face recognition and memory, and greater brain atrophy has been observed among patients in the early stages of AD. Accordingly, the proposed GCN-EASA module is capable of identifying subject-specific pathological regions.

B. Identification of Risk SNP Loci Associated With MCI

The identification of risk SNP loci is crucial for computer-assisted early diagnosis of AD. For this purpose, the MA-DFN approach was utilized to investigate risk SNP loci with potential biological significance. These loci have been identified as biomarkers that impact early neurological disorders. In this study, the weights of the first layer in the DAD-AE module are crucial for identifying key SNPs that are relevant to the final classification task. To obtain reliable results, we used 10-fold cross-validation to average the weights of the first layer. Subsequently, we selected the SNP loci corresponding to the top-15 weights as potential disease-related biomarkers. Table III presents a summary of the top-15 risk SNP loci, which have the highest degree of differentiation and influence for the CN versus MCI task. The table comprises 4 columns with information including SNPs IDs, risk alleles, overlapping or closest genes, and chromosome (CHR).

Table III shows that our proposed method has successfully identified the top-15 SNPs significantly associated with MCI progression. Among them, APP and APOE have been identified as being associated with the developmental process of MCI [56]. BIN1, ABCA7, and SLC4A4 were identified as the major susceptibility genes for MCI [57]. Therein, BIN1 contributes to the deregulation of early endosomal size, which is a hallmark of early MCI pathophysiology. These risk SNP loci pertaining to MCI exemplify the proposed algorithm's capability to enhance the diagnosis of brain diseases by effectively characterizing the data.

VI. CONCLUSION

This work developed an MA-DFN for brain imaging genomic-assisted diagnosis. The critical distinction between the work presented here and that proposed in recent multimodal deep learning frameworks is the combination of intra-modal geometric properties, inter-modal local discriminability, and consistency. Multimodal brain imaging genomics datasets are often complex, thus raising the critical question of how to faithfully describe the true structure of the data for classification purposes. One can easily apply the standard GCNs to brain network data without considering the node-structure feature relationships and long-range dependencies between brain regions. To address this, the proposed GCN-EASA module effectively captures long-range dependencies between brain regions and extracts valuable structural information from brain networks to support downstream tasks. Additionally, a DAD-AE has been introduced to capture the distribution of SNP data and mitigate the impact of irrelevant noise on the identification of important gene loci. Moreover, the MA-DFN approach uniquely maximizes the utility of multimodal data, unlike existing joint imaging genomics multimodal diagnostic methods that fail to simultaneously exploit the consistency and discriminatory nature of multimodality. The MA-DFN method achieves this by incorporating a DFN-CorrReg module, which projects multimodal features into a common space. In this common space, intra-class and inter-modal samples are expected to be both compact and correlated, while between-class samples should be distinguishable. The experimental results demonstrate that the proposed method's accuracy on ADNI imaging genomics data surpasses current state-of-the-art methods.

REFERENCES

[1] J. Lee, "Mild cognitive impairment in relation to Alzheimer's disease: An investigation of principles, classifications, ethics, and problems," *Neuroethics*, vol. 16, no. 2, p. 16, Jul. 2023.
[2] B. B. Patel and N. W. Holland, "Mild cognitive impairment: Hope for stability, plan for progression," *Cleveland Clinic J. Med.*, vol. 79, no. 12, pp. 857–864, Dec. 2012.
[3] H. T. Gorji and N. Kaabouch, "A deep learning approach for diagnosis of mild cognitive impairment based on MRI images," *Brain Sci.*, vol. 9, no. 9, p. 217, Aug. 2019.
[4] C. Platero and M. C. Tobar, "Predicting Alzheimer's conversion in mild cognitive impairment patients using longitudinal neuroimaging and clinical markers," *Brain Imag. Behav.*, vol. 15, no. 4, pp. 1728–1738, Aug. 2021.
[5] R. C. Petersen, B. Caracciolo, C. Brayne, S. Gauthier, V. Jelic, and L. Fratiglioni, "Mild cognitive impairment: A concept in evolution," *J. Internal Med.*, vol. 275, no. 3, pp. 214–228, Mar. 2014.
[6] F. Jacini et al., "Amnestic mild cognitive impairment is associated with frequency-specific brain network alterations in temporal poles," *Frontiers Aging Neurosci.*, vol. 10, p. 400, Dec. 2018.
[7] J. Chen, D. Yang, H. Cai, M. Styner, and G. Wu, "Discovering spreading pathways of neuropathological events in Alzheimer's disease using harmonic wavelets," in *Proc. 27th Int. Conf. Inf. Process. Med. Imag. (IPMI)*. Bornholm, Denmark: Springer, Jun. 2021, pp. 228–240.
[8] C. Gaiteri, S. Mostafavi, C. J. Honey, P. L. De Jager, and D. A. Bennett, "Genetic variants in Alzheimer disease-molecular and brain network approaches," *Nature Rev. Neurol.*, vol. 12, no. 7, pp. 413–427, Jul. 2016.
[9] S. S. M. Elsheikh, E. R. Chimusa, N. J. Mulder, and A. Crimi, "Genome-wide association study of brain connectivity changes for Alzheimer's disease," *Sci. Rep.*, vol. 10, no. 1, p. 1433, 2020.
[10] D. Zhang, J. Huang, B. Jie, J. Du, L. Tu, and M. Liu, "Ordinal pattern: A new descriptor for brain connectivity networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1711–1722, Jul. 2018.
[11] C. Lian, M. Liu, L. Wang, and D. Shen, "Multi-task weakly-supervised attention network for dementia status estimation with structural MRI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 4056–4068, Aug. 2022.

[12] J. Ji, A. Zou, J. Liu, C. Yang, X. Zhang, and Y. Song, "A survey on brain effective connectivity network learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 1879–1899, Apr. 2023.

[13] X.-A. Bi et al., "Community graph convolution neural network for Alzheimer's disease classification and pathogenetic factors identification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 19, 2023, doi: 10.1109/TNNLS.2023.3269446.

[14] A. Sharma and P. Dey, "A machine learning approach to unmask novel gene signatures and prediction of Alzheimer's disease within different brain regions," *Genomics*, vol. 113, no. 4, pp. 1778–1789, Jul. 2021.

[15] W. Hu et al., "Interpretable multimodal fusion networks reveal mechanisms of brain cognition," *IEEE Trans. Med. Imag.*, vol. 40, no. 5, pp. 1474–1483, May 2021.

[16] A. Gossmann, P. Zille, V. Calhoun, and Y.-P. Wang, "FDR-corrected sparse canonical correlation analysis with applications to imaging genomics," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1761–1774, Aug. 2018.

[17] J. Ansarifar, I. I. Karipidis, M. Saggar, and D. S. Hong. (2023). *BIG-graph: Brain Imaging Genetics By Graph Neural Network*. [Online]. Available: https://openreview.net/forum?id=jxAle6zdyTI

[18] R. Guo et al., "Graph-based fusion of imaging, genetic and clinical data for degenerative disease diagnosis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 21, no. 1, pp. 57–68, Feb. 2024.

[19] L. Hernández-Lorenzo, M. Hoffmann, E. Scheibling, M. List, J. A. Matías-Guiu, and J. L. Ayala, "On the limits of graph neural networks for the early diagnosis of Alzheimer's disease," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, Oct. 2022.

[20] Y. Huang and A. C. Chung, "Disease prediction with edge-variational graph convolutional networks," *Med. Image Anal.*, vol. 77, Apr. 2022, Art. no. 102375.

[21] J. Liu, H. Du, R. Guo, H. X. Bai, H. Kuang, and J. Wang, "MMGK: Multimodality multiview graph representations and knowledge embedding for mild cognitive impairment diagnosis," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 1, pp. 389–398, Feb. 2024.

[22] C. J. Stam, "Modern network science of neurological disorders," *Nature Rev. Neurosci.*, vol. 15, no. 10, pp. 683–695, Oct. 2014.

[23] K. Wang, Z. Li, Z.-H. You, P. Han, and R. Nie, "Adversarial dense graph convolutional networks for single-cell classification," *Bioinformatics*, vol. 39, no. 2, Feb. 2023, Art. no. btad043.

[24] V. P. Dwivedi et al., "A generalization of transformer networks to graphs," in *Proc. AAAI Workshop Deep Learn. Graphs, Methods Appl.*, 2021, pp. 1–8.

[25] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, "Brain network transformer," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 25586–25599.

[26] P. Xu, X. Zhu, and D. A. Clifton, "Multimodal learning with transformers: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12113–12132, May 2023.

[27] S. Ghosal et al., "G-MIND: An end-to-end multimodal imaging-genetics framework for biomarker identification and disease classification," *Proc. SPIE*, vol. 11596, pp. 63–72, Feb. 2021.

[28] J. X. Wang, Y. Li, X. Li, and Z.-H. Lu, "Alzheimer's disease classification through imaging genetic data with IGnet," *Frontiers Neurosci.*, vol. 16, Mar. 2022, Art. no. 846638.

[29] W. Ko, W. Jung, E. Jeon, and H.-I. Suk, "A deep generative–discriminative learning for multimodal representation in imaging genetics," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2348–2359, Sep. 2022.

[30] M. A. Rahaman et al., "Deep multimodal predictome for studying mental disorders," *Hum. Brain Mapping*, vol. 44, no. 2, pp. 509–522, Feb. 2023.

[31] L. Du et al., "Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101656.

[32] Y. Yu, S. Tang, K. Aizawa, and A. Aizawa, "Category-based deep CCA for fine-grained venue discovery from multimodal data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1250–1258, Apr. 2019.

[33] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.

[34] K. Jia, J. Lin, M. Tan, and D. Tao, "Deep multi-view learning using neuron-wise correlation-maximizing regularizers," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5121–5134, Oct. 2019.

[35] F. Tramer, D. Boneh, A. Kurakin, I. Goodfellow, N. Papernot, and P. McDaniel, "Ensemble adversarial training: Attacks and defenses," in *Proc. Int. Conf. Learn. Represent.*, vol. 1, 2018, p. 2.

[36] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 21–25.

[37] J. Ansarifar and L. Wang, "New algorithms for detecting multi-effect and multi-way epistatic interactions," *Bioinformatics*, vol. 35, no. 24, pp. 5078–5085, Dec. 2019.

[38] X. Li, Q. Wang, F. Nie, and M. Chen, "Locality adaptive discriminant analysis framework," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7291–7302, Aug. 2022.

[39] P. Hu, D. Peng, Y. Sang, and Y. Xiang, "Multi-view linear discriminant analysis network," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5352–5365, Nov. 2019.

[40] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.

[41] N. Toussaint, J.-C. Souplet, and P. Fillard, "MedINRIA: Medical image navigation and research tool by INRIA," in *Proc. MICCAI Workshop Interact. Med. Image Anal. Vis.*, 2007, pp. 1–8.

[42] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012.

[43] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, "Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature," *NeuroImage*, vol. 53, no. 1, pp. 1–15, Oct. 2010.

[44] S. Purcell et al., "PLINK: A tool set for whole-genome association and population-based linkage analyses," *Amer. J. Human Genet.*, vol. 81, no. 3, pp. 559–575, Sep. 2007.

[45] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.

[46] H. Wang, F. Nie, H. Huang, S. L. Risacher, A. J. Saykin, and L. Shen, "Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," *Bioinformatics*, vol. 28, no. 12, pp. i127–i136, Jun. 2012.

[47] K. Wei, W. Kong, and S. Wang, "An improved multi-task sparse canonical correlation analysis of imaging genetics for detecting biomarkers of Alzheimer's disease," *IEEE Access*, vol. 9, pp. 30528–30538, 2021.

[48] T. Zhou, K. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Hum. Brain Mapping*, vol. 40, no. 3, pp. 1001–1016, Feb. 2019.

[49] G. Li et al., "Application of deep canonically correlated sparse autoencoder for the classification of schizophrenia," *Comput. Methods Programs Biomed.*, vol. 183, Jan. 2020, Art. no. 105073.

[50] W. Hu, B. Cai, A. Zhang, V. D. Calhoun, and Y.-P. Wang, "Deep collaborative learning with application to the study of multimodal brain development," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 12, pp. 3346–3359, Dec. 2019.

[51] S. Moon, J. Hwang, and H. Lee, "SDGCCA: Supervised deep generalized canonical correlation analysis for multi-omics integration," *J. Comput. Biol.*, vol. 29, no. 8, pp. 892–907, Aug. 2022.

[52] M. Wang, W. Shao, S. Huang, and D. Zhang, "Deep self-reconstruction sparse canonical correlation analysis for brain imaging genetics," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 1790–1793.

[53] T. Gómez-Isla et al., "Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease," *Ann. Neurol.*, vol. 41, no. 1, pp. 17–24, Jan. 1997.

[54] D. Vatansever, A. E. Manktelow, B. J. Sahakian, D. K. Menon, and E. A. Stamatakis, "Angular default mode network connectivity across working memory load," *Hum. Brain Mapping*, vol. 38, no. 1, pp. 41–52, Jan. 2017.

[55] O. Woolnough, P. S. Rollo, K. J. Forseth, C. M. Kadipasaoglu, A. D. Ekstrom, and N. Tandon, "Category selectivity for face and scene recognition in human medial parietal cortex," *Current Biol.*, vol. 30, no. 14, pp. 2707–2715, Jul. 2020.

[56] R. J. Guerreiro, D. R. Gustafson, and J. Hardy, "The genetic architecture of Alzheimer's disease: Beyond APP, PSENs and APOE," *Neurobiol. Aging*, vol. 33, no. 3, pp. 437–456, Mar. 2012.

[57] L. Yu et al., "Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease," *J. Amer. Med. Assoc. Neurol.*, vol. 72, no. 1, pp. 15–24, 2015.

Xiaoqi Sheng received the B.S. and M.S. degrees from the School of Information Science and Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China.

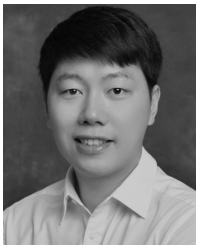
His current research interests include medical imaging, brain network analysis, and machine learning.

Hongmin Cai (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics from Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, and the Ph.D. degree in applied mathematics from The University of Hong Kong, Hong Kong, in 2007.

He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include biomedical image processing and omics data integration.

Yongwei Nie (Member, IEEE) received the B.Sc. and Ph.D. degrees from the Computer School, Wuhan University, Wuhan, China, in 2009 and 2015, respectively.

He is currently an Associate Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include image and video editing and computational vision.



Shengfeng He (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Macau University of Science and Technology, Macau, China, in 2009 and 2011, respectively, and the Ph.D. degree from City University of Hong Kong, Hong Kong, in 2015.

He was a Faculty Member with South China University of Technology, Guangzhou, China, from 2016 to 2022. He is currently an Associate Professor with the School of Computing and Information Systems, Singapore Management University, Singapore. His research interests include computer vision and generative models.

Dr. He is a Senior Member of CCF. He serves as the Lead Guest Editor for *International Journal of Computer Vision* and an Associate Editor for *Neurocomputing*.

Yiu-Ming Cheung (Fellow, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2000.

He is currently the Chair Professor of the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His current research interests include machine learning, pattern recognition, visual computing, and their application.

Dr. Cheung is a fellow of AAAS, IET, and BCS. He is the Founding Chairperson of the Computational Intelligence Chapter of the IEEE Hong Kong Section. He is the Editor-in-Chief of IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE. He serves as an Associate Editor for IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (2014–2020), *Pattern Recognition*, *Knowledge and Information Systems*, and *Neurocomputing*.

Jiazhou Chen (Member, IEEE) received the Ph.D. degree in computer science and technology from South China University of Technology (SCUT), Guangzhou, China, in 2020.

During his period at SCUT, he studied in the Department of Psychiatry, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, as a Visiting Scholar, in 2019. He is currently an Associate Professor with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou. His current research interests include bioinformatics, medical image analysis, brain network analysis, and machine learning.